

This is the peer-reviewed version of the following article:

Peón, A., Coderch, C., Gago, F., & González-Bello, C. (2013). Comparative Binding Energy (COMBINE) Analysis for Understanding the Binding Determinants of Type II Dehydroquinase Inhibitors. *Chemmedchem*, 8(5), 740-747, which has been published in final form at <https://doi.org/10.1002/cmdc.201300013>.

This article may be used for non-commercial purposes in accordance with Wiley-VCH Terms and Conditions for Self-Archiving.

DOI: 10.1002/cmdc.200((will be filled in by the editorial staff))

# Comparative Binding Energy (COMBINE) Analysis for Understanding the Binding Determinants of Type II Dehydroquinase Inhibitors

Antonio Peón,<sup>[a]</sup> Claire Coderch,<sup>[b]</sup> Federico Gago,<sup>[b]</sup> and Concepción González-Bello<sup>\*[a]</sup>

*A COMBINE analysis to derive QSAR models that help rationalize the determinants of binding affinity for inhibitors of the third enzyme of the shikimic acid pathway, type II dehydroquinase (DHQ2), is reported. Independent COMBINE models were derived for Helicobacter pylori and Mycobacterium tuberculosis DHQ2, which is an essential enzyme in these pathogenic bacteria that has no counterpart in human cells.*

*These studies quantify the importance of the hydrogen bonding interaction of the ligands with the water molecule involved in the enzymatic mechanism and highlight important differences in the ligand interactions with the interface pocket close to the active site that could provide guides for future inhibitor designs.*

## Introduction

The shikimic acid pathway that is present in bacteria, fungi, plants, and in certain apicomplexan parasites, but it is absent in mammals, is considered an attractive target for the development of new antimicrobials and herbicides.<sup>[1,2]</sup> In recent years a great deal of effort has been focused on the development of inhibitors of the enzymes involved in this route. Among them, the third enzyme in the pathway, the type II dehydroquinase (3-dehydroquinase dehydratase, EC 4.2.1.10, DHQ2), has generated particular interest because it is essential in important pathogenic bacteria such as, *Mycobacterium tuberculosis*, the causative agent of tuberculosis, and *Helicobacter pylori*, the causative agent of gastric and duodenal ulcers, which has also been classified as a type I carcinogen.<sup>[3]</sup>

DHQ2 catalyzes the reversible dehydration of 3-dehydroquinic acid (**1**) to form 3-dehydroshikimic acid (**2**) (Scheme 1). The enzymatic conversion is initiated by an essential tyrosine of the active site that removes the pro-S hydrogen from C2 of **1**.<sup>[4,5]</sup> The elimination proceeds through a stepwise E<sub>1</sub>CB mechanism<sup>[4]</sup> involving an enol intermediate **3**, which is stabilized by a conserved water molecule. The final step is the acid-catalyzed elimination of the C1 hydroxyl group – a reaction mediated by a conserved histidine acting as a proton donor.

As a result of these investigations, a large number of competitive reversible inhibitors have been developed that include analogs of the natural substrate, e.g. compounds **4**,<sup>[6]</sup> and mimetics of the intermediate of the enzyme-catalyzed reaction, e.g. compounds **5**<sup>[7-19]</sup> and the bicyclic derivatives **6**<sup>[20]</sup> (Figure 1).<sup>[21,22]</sup> In addition, crystal structures of several enzyme-inhibitor binary complexes with both types of inhibitors have also been described.<sup>[7-9,20,23-25]</sup>

The resolution of these crystal structures has been very important to understand the role of the substituents <sup>1</sup>R–<sup>4</sup>R in compounds **4**–

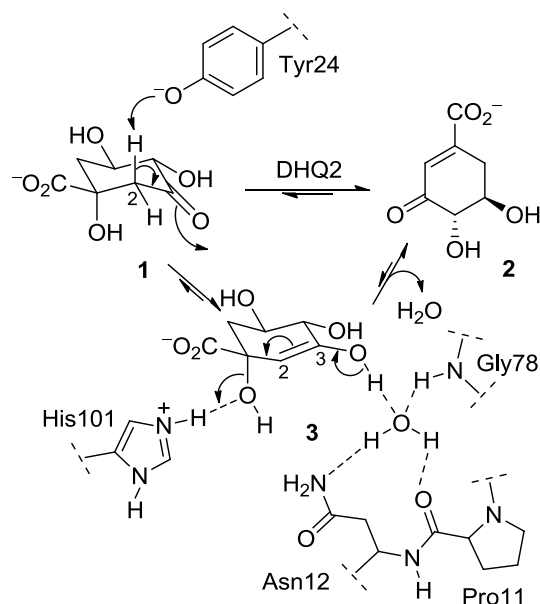
**6**, in particular, their interaction with key residues responsible for catalysis, both located in the flexible loop that closes over the active site upon substrate binding. Besides, a detailed knowledge of the contribution of each active site residue to the observed inhibition activity would provide valuable information that could help improve our understanding of some of the determinants of binding affinity and guide further inhibitor design. For this purpose, COMparative BINDing Energy (COMBINE) analysis, a chemometric method that relies on a number of three-dimensional complexes to calculate, decompose and weigh residue-based ligand-receptor interaction energies (both van der Waals and electrostatic),<sup>[26]</sup> has become particularly useful in the study of quantitative structure-activity relationships (QSAR). It has been successfully applied to protein-ligand,<sup>[27-31]</sup> RNA-ligand,<sup>[32]</sup> protein-DNA<sup>[33]</sup> and protein-peptide complexes.<sup>[34]</sup> The data matrix generated by following this approach is then subjected to multivariate statistical analysis to try and correlate the computed energy components with experimental measures of activity or affinity using Partial Least Squares (PLS) analysis.

In the following we show how COMBINE analysis was useful to

- [a] A. Peón and Prof. C. González-Bello  
Centro Singular de Investigación en Química Biológica y Materiales Moleculares (CIQUS),  
Universidad de Santiago de Compostela,  
calle Jenaro de la Fuente s/n, 15782 Santiago de Compostela,  
Spain  
Fax: (+34) 881 815704  
E-mail: [concepcion.gonzalez.bello@usc.es](mailto:concepcion.gonzalez.bello@usc.es)
- [b] Dr. C. Coderch and Prof. F. Gago  
Departamento de Ciencias Biomédicas II,  
Universidad de Alcalá,  
28871 Alcalá de Henares, Madrid, Spain

Supporting information for this article is available on the WWW under <http://www.chemmedchem.org> or from the author

gain critical information regarding the interactions in the active site of a series of competitive reversible inhibitors of DHQ2 enzymes from *H. pylori* (DHQ2-Hp) and *M. tuberculosis* (DHQ2-Mt) and derive QSAR models that aid in the rationalization of the determinants of binding affinity and in understanding the differences between both enzymes.



Scheme 1. Enzymatic conversion of 3-dehydroquinic acid (**1**) to 3-dehydroshikimic acid (**2**) catalyzed by DHQ2. The reaction proceeds via an enol intermediate **3**. Relevant residues are indicated (the numbering corresponds to *M. tuberculosis*).

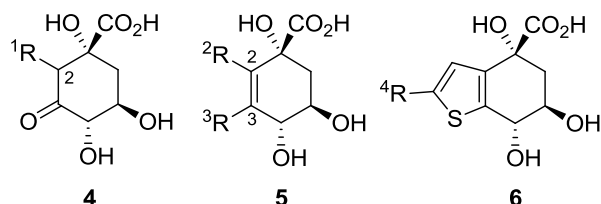


Figure 1. Competitive reversible inhibitors of DHQ2. 2-alkyl ketones **4** are substrate analogs and compounds **5** and **6** are mimetics of the intermediate of the enzyme-catalyzed reaction.

## Results and Discussion

The COMBINE analysis with the *H. pylori* enzyme was carried out on 54 enzyme-inhibitor complexes modeled after eight DHQ2-Hp/ligand co-crystal structures (PDB accession codes: 2WKS,<sup>[7]</sup> 1J2Y,<sup>[35]</sup> 2C4W,<sup>[24]</sup> 2C57,<sup>[24]</sup> 2XB9,<sup>[23]</sup> 2XDA,<sup>[20]</sup> 2C4V,<sup>[24]</sup> and 2XD9<sup>[20]</sup>). The inhibitors covered an activity range, expressed as  $pK_i$ , between 3.4 and 7.0. The best cross-validated PLS model had four latent variables (supporting information). For the DHQ2-Mt enzyme, 63 compounds covering an activity range between 3.7 and 7.6  $pK_i$  units, and 13 crystallographic binary DHQ2-Mt/ligand complexes (PDB accession codes: 1H0R,<sup>[36]</sup> 1H0S,<sup>[36]</sup> 2Y71,<sup>[8]</sup> 2Y76,<sup>[8]</sup> 2Y77,<sup>[8]</sup> 2XB8,<sup>[23]</sup> 3N59,<sup>[9]</sup> 3N8K,<sup>[9]</sup> 3N76,<sup>[9]</sup> 3N87,<sup>[9]</sup> 3N8N,<sup>[9]</sup> 3N7A<sup>[9]</sup> and 1H05<sup>[37]</sup>) were employed. For this enzyme, the best cross-validated PLS model has six latent variables (supporting information). In both studies, the  $pK_i$  values proved to be highly correlated with a set of weighted ligand-

residue interaction energies ( $r^2 = 0.83 \pm 0.07$  and  $0.88 \pm 0.05$ , respectively, in fitting and  $q^2 = 0.73 \pm 0.11$  and  $0.71 \pm 0.11$ , respectively, in cross-validation; Figure 2).

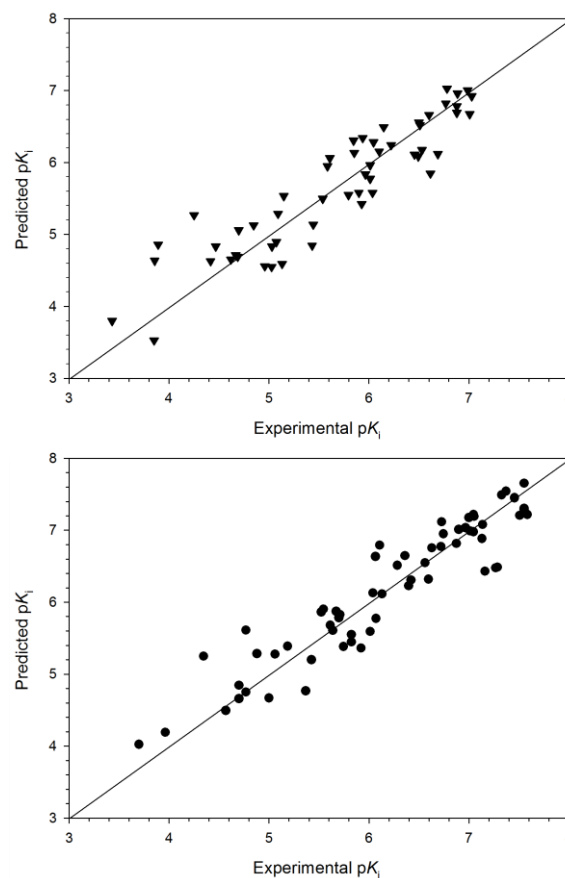


Figure 2. Correlation between experimental  $pK_i$  values and those calculated using the COMBINE QSAR models for DHQ2-Hp (top, 4 latent variables) and DHQ2-Mt (bottom, 6 latent variables).

The PLS pseudo-coefficients (hereafter coefficients) for the calculated van der Waals and electrostatic interactions corresponding to the best model of each enzyme are shown in Figure 3 and summarized in Table 1.

The sign and the magnitude of the PLS coefficient assigned to an energy term multiplied by the corresponding energy value reveals the relative importance of each residue in accounting for the observed differences in binding affinity. Since inhibition constants (i.e.  $K_i$  or dissociation constants of the enzyme-inhibitor complexes) can be directly related to the free energy of binding to the enzyme as  $\Delta G_{\text{binding}} = -2.303 RT pK_i$ , the chemometric meaning of the values obtained is that the larger a negative PLS coefficient multiplying a favorable interaction energy (negative value), the greater the  $pK_i$  and vice versa.

In general, the COMBINE analysis for the DHQ2-Hp enzyme identifies that the electrostatic and hydrogen bonding interaction with the essential water molecule (WAT<sub>e</sub>), is the most significant contribution to the inhibition potency. This study also shows that van der Waals interactions involving the active site residues Asn10, Met13, Leu14, Tyr22, Leu103, His102 and Gly78, together with residues Gln63\* and Ala90\* from a symmetry-related neighboring enzyme subunit, are those that contribute most significantly to explaining the differences in inhibitory potency as these residues have the largest PLS coefficients

assigned to their van der Waals interactions with the ligands. For DHQ2-Mt, the COMBINE analysis suggests that the potency of the inhibitors is enhanced by promoting van der Waals interactions with active site residues Tyr24, Leu13, Leu16, Ile102, Gly77, Asn75 and His101, and also with residues Ala91\* and Glu92\* from a neighboring subunit, as well as by electrostatic

and/or hydrogen bonding interactions with the essential water molecule (WAT<sub>E</sub>) and with a crystallographic water molecule (WAT<sub>C</sub>) located near Arg108 and Tyr24. In both cases, these key interactions can be organized into five key pockets as follows (Figure 4):

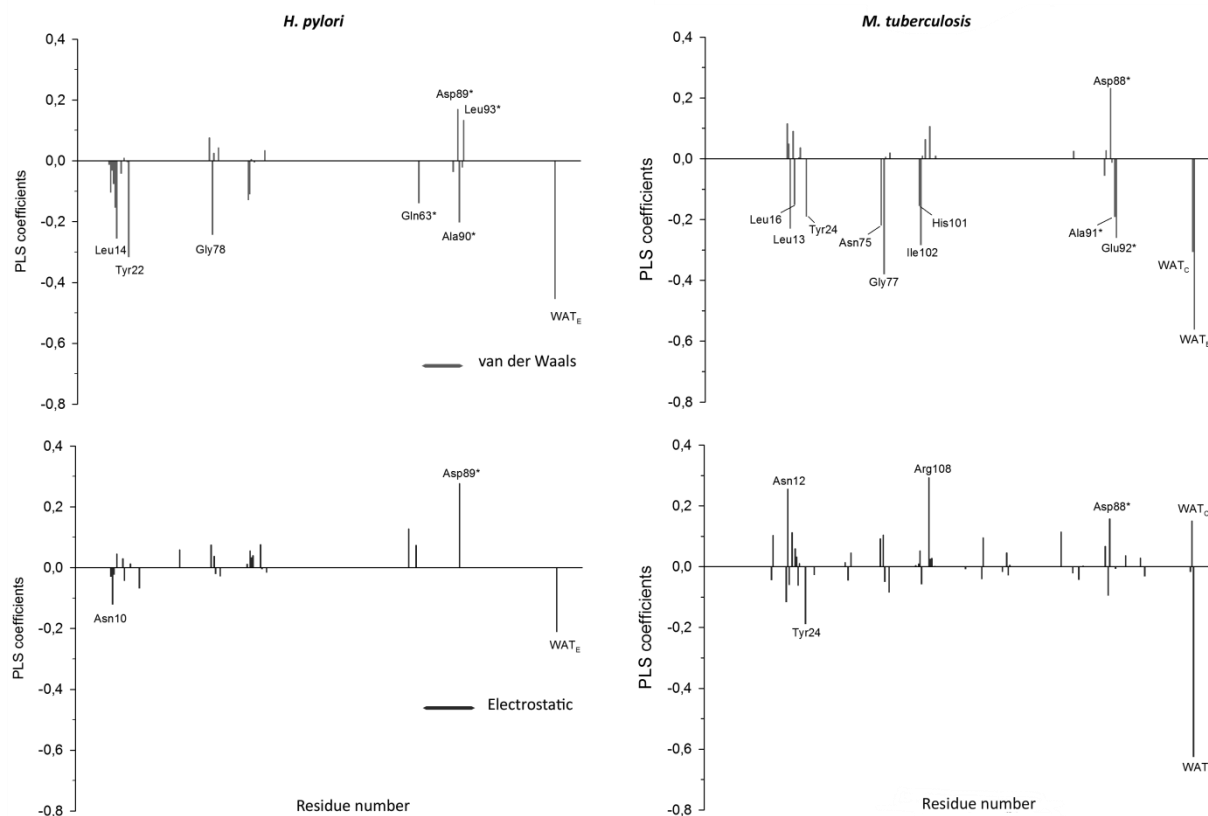


Figure 3. Plot of normalized PLS coefficients of the van der Waals (top) and electrostatic (bottom) interactions in the COMBINE models for DHQ2-Hp (left, 4 latent variables) and DHQ2-Mt (right, 6 latent variables). The residues with the highest coefficients have been labeled. WAT<sub>E</sub> and WAT<sub>C</sub> stand for the essential and crystallographic water molecules, respectively.

**Loop – Essential tyrosine.** The COMBINE analysis for DHQ2-Hp reveals that the interaction with Tyr22, which has a PLS coefficient of -0.32, is one of the most relevant interactions for enhancing inhibition of this enzyme. Tyr22, as well as Arg17, has been identified by chemical modification and site-directed mutagenesis studies as being essential for enzyme activity.<sup>[42,43]</sup> Both residues are located in the flexible loop that closes over the active site upon substrate binding. Molecular dynamics (MD) simulation studies suggested that the side chain of Tyr22 must be perpendicular to the cyclohexane ring of the natural substrate for proton abstraction.<sup>[23,25]</sup> The favorable interaction of the aromatic rings of the ligands with the Tyr22 side chain prevents its appropriate orientation for catalysis. Such favorable interaction causes a drastic reduction of the loop flexibility that is essential for enzyme function. Comparison of the PLS coefficients for the essential tyrosine obtained for both enzymes shows that the van der Waals interaction with this residue is less significant for DHQ2-Mt. We consider that these differences are due to the distinct loop flexibility and motion in the two enzymes which we have studied in detail.<sup>[25]</sup> MD simulation results suggest that the essential arginine of this loop has a crucial role in controlling the position and appropriate orientation of Tyr22 whereas in DHQ2-Mt this tyrosine side chain motion is not controlled by the

essential arginine. Instead, its position could be mainly controlled by the interaction with the aromatic moiety of the ligand. These differences make DHQ2-Hp particularly sensitive to changes introduced at the C2 position of the natural substrate because the loop should be less able to accommodate these changes than the equivalent loop in DHQ2-Mt. Moreover, for this latter enzyme, the COMBINE analysis reveals an unfavorable contribution of the electrostatic interaction with Arg108 (a PLS coefficient of +0.30). This residue has a key role in the function of the Tyr24 residue. When the substrate binds to the enzyme, the loop is closed over and a strong hydrogen bonding interaction between the phenolic OH of Tyr24 and the conserved Arg108 is formed. Regardless of the appropriate (catalysis) or inappropriate (inhibition) conformation of the tyrosine side chain, the hydrogen bonding interaction between both residues is crucial. Hence, this COMBINE study predicts that those ligands having a favorable electrostatic interaction with Arg108, for instance through the C1 carboxylate group, will show lower affinity for the enzyme. The fact that a similar situation was not predicted for DHQ2-Hp can be explained by considering once again the differences in loop motion and flexibility of both enzymes. In fact, in all the available DHQ2-Hp crystal structures, the essential tyrosine's side chain establishes a strong hydrogen bonding interaction with the

guanidinium group of Arg109. In contrast, for DHQ2-Mt, this hydrogen bonding interaction is not always observed, and in some cases, the position of this residue has not been located in the electron density map.

**Table 1.** Selection of the PLS pseudo-coefficients assigned to the van der Waals and electrostatic interaction energies obtained from the COMBINE analyses of DHQ2-Hp and DHQ2-Mt.<sup>[a]</sup>

<i>H. pylori</i>			<i>M. tuberculosis</i>		
Residue	vdW	Elec	Residue	vdW	Elec
Tyr22	-0.32	--	Tyr24	-0.19	-0.19
Met13	-0.15	--	Arg108	--	0.30
Gln63*	-0.14	--			
Asp89*	0.17	0.28	Asp88*	0.23	0.16
Ala90*	-0.20	--	Ala91*	-0.21	--
Leu93*	0.13	--	Glu92*	-0.26	--
Leu14	-0.26	--	Leu13	-0.23	--
			Leu16	-0.15	--
Leu103	-0.11	--	Ile102	-0.28	--
Gly78	-0.24	--	Gly77	-0.38	--
			Pro11	0.12	-0.14
Asn10	--	-0.12	Asn12	--	0.26
			Asn75	-0.22	--
His102	-0.13		His101	-0.16	--
WAT <sub>E</sub>	-0.45	-0.21	WAT <sub>E</sub>	-0.56	-0.62
			WAT <sub>C</sub>	-0.31	0.15

[a] Only PLS pseudo-coefficients with absolute values greater than 0.1 are displayed. WAT<sub>E</sub> and WAT<sub>C</sub> are the essential and crystallographic water molecules, respectively.

**Interface between chains.** This is a very important pocket<sup>[9]</sup> of the enzyme that reveals significant differences between both enzymes and whose contribution can be now quantified. For DHQ2-Hp, the results indicate that the van der Waals or hydrogen bonding interactions of the ligands with Met13 and two residues of a symmetry-related neighboring subunit, Gln63\* and Ala90\*, would improve their inhibition potency. For the latter residues, negative PLS pseudo-coefficients of -0.15, -0.14 and -0.20, were assigned (Figure 4a). On the contrary, van der Waals interactions with residues Leu93\* and Asp89\*, also located in the neighboring subunit, are identified as unfavorable, showing positive PLS pseudo-coefficients of 0.13 and 0.17, respectively. For DHQ2-Mt, a negative coefficient for the favorable van der Waals interaction, which is quantitatively similar to that assigned to Tyr24, with residues Ala91\* and Glu92\* located in the neighboring subunit is identified (Figure 4c). For the latter residues, negative PLS pseudo-coefficients of -0.21 and -0.26, respectively, were obtained.

A comparison of the amino acid sequences of both enzymes reveals that there are important differences in the type of residues located in this pocket (see supporting information). The pocket is more polar in DHQ2-Mt than in DHQ2-Hp as it is made up of His63\*, Glu92\*, Asp67\*, Arg15, Asp53, and Asn12 residues in the former and Ile59\*, Gln63\*, Leu93\*, Val66\*, Met13, Asn53, Asn10 and Asn12 in the latter. This explains the fact that, in some cases, the same ligand shows marked differences in its interaction with

both enzymes. These COMBINE analyses quantify these differences and provide guides for further inhibitor design. Thus, for DHQ2-Hp, these results suggest that those compounds functionalized with a flexible chain located at the bottom of the interface might provide more potent inhibitors, since this would enhance their interaction with residues Gln63\* and Ala90\* and avoid contacts with Asp89\* and Leu93\* which are the residues located at the top of the interface that this COMBINE study identifies as unfavorable. For DHQ2-Mt, the results suggest that the inhibitory potency of the ligands might be enhanced by establishing favorable van der Waals interactions with Glu92\*, which is located at the entrance to the interface between two neighboring subunits and not with the side chain residues at the bottom of the interface as in DHQ2-Hp. In common with the analysis on DHQ2-Hp, COMBINE identifies any contacts with the Asp88\* located in the neighboring subunit as unfavorable. For both enzymes, this conserved residue is involved in the recognition of the hydroxyl group at the C4 position in both the natural substrate and the reaction intermediate.

**C2 binding pocket.** This is a key region of the active site for ligand binding. For DHQ2-Hp, it is lined by residues Leu14, Gly78 and Leu103, which are assigned negative PLS coefficients of -0.26, -0.24 and -0.11, respectively (Figure 4b). For DHQ2-Mt, in addition to the equivalent residues, Leu13, Gly77 and Ile102, Leu16 is also identified and negative PLS coefficients of -0.23, -0.38, -0.28 and -0.15, respectively, are assigned to them (Figure 4d). In general, the contribution of van der Waals interactions with these residues to the inhibitory potencies is greater than those corresponding to residues at the interface pocket, as reflected by the PLS coefficients obtained, in particular those related to Leu14 and Gly78 for DHQ2-Hp, and Ile102 and Gly77 for DHQ2-Mt. These results justify the high inhibitory potency of the natural substrate analogs substituted at C2 with a benzyl group in the equatorial position.<sup>[6]</sup> It is worth mentioning that these compounds still contain the axial hydrogen that could be removed by the essential tyrosine. The PLS coefficients assigned to Ile102 (-0.28) in DHQ2-Mt is more than twice the value assigned to the equivalent Leu103 in DHQ2-Hp (-0.11). In principle, the structural differences between the two enzymes in this region of the active site are small and do not justify the observed differences. However, because this apolar pocket is generated as a result of the closure of the loop upon substrate or inhibitor binding, the distinct flexibility and motion of the loops in the two enzymes are likely to account for these differences.

**Essential water molecule.** Our results show that the electrostatic and hydrogen bonding interactions of the inhibitors with the water molecule involved in the enzymatic mechanism, which orients the natural substrate and the intermediate in the active site, contribute favorably to potency against both enzymes. In fact, this is the van der Waals interaction (hydrogen bonding) with the highest negative PLS coefficients values (-0.45 and -0.56, for DHQ2-Hp and DHQ2-Mt enzymes, respectively). These results quantify the experimental observation that the replacement of the oxygen atom of the 3-methoxyaryl enol mimetics by a carbon atom leads to a decrease in the inhibition potency of up to 20-fold (supporting information, Table S1, entry 1).<sup>[17]</sup> Our chemometric studies quantitate how important this contact is for modulating the potency of the inhibitors and reveal that it is worth considering in future medicinal chemistry designs.

On the other hand, for DHQ2-Mt, a favorable hydrogen bonding interaction with a water molecule observed in several crystal structures is identified. This water molecule is located very close to the essential Tyr24 and the conserved Arg108. For instance, in PDB entry 2Y71<sup>[8]</sup> solved at 1.5 Å, this water molecule is 2.5 Å

and 2.9 Å away from Tyr24 and Arg108, respectively. We consider that this result reveals once again the importance of the inappropriate orientation of Tyr24 for proton abstraction and therefore for good enzyme inhibition.

**C1 binding pocket.** These COMBINE analyses reveal that the hydrogen bonding interaction between the ligands and the essential histidine, His102 for DHQ2-Hp and His101 for DHQ2-Mt, is also important for modulating the inhibitory potency against both enzymes, as attested by the negative PLS coefficients of  $-0.13$  and  $-0.16$ , respectively. This residue is involved in the final elimination step by acting as the general acid with a conserved asparagine, Asn76 for DHQ2-Hp and Asn75 for DHQ2-Mt, and

correctly positioning the C1 hydroxyl group to accept a proton. For the latter enzyme, a negative PLS coefficient value of  $-0.22$  was also obtained for the favorable hydrogen-bonding interaction with Asn75. For a long time, it has been suggested that the C1 binding pocket plays a crucial role in ligand recognition. In fact, it is believed that the driving force for product release during catalysis is the conformational change in the product as the C1 atom becomes planar.<sup>[5]</sup> The present COMBINE analyses show that DHQ2-Mt is particularly sensitive to small conformational changes in the inhibitor because they may change the correct positioning of the C1 hydroxyl group for hydrogen bonding.

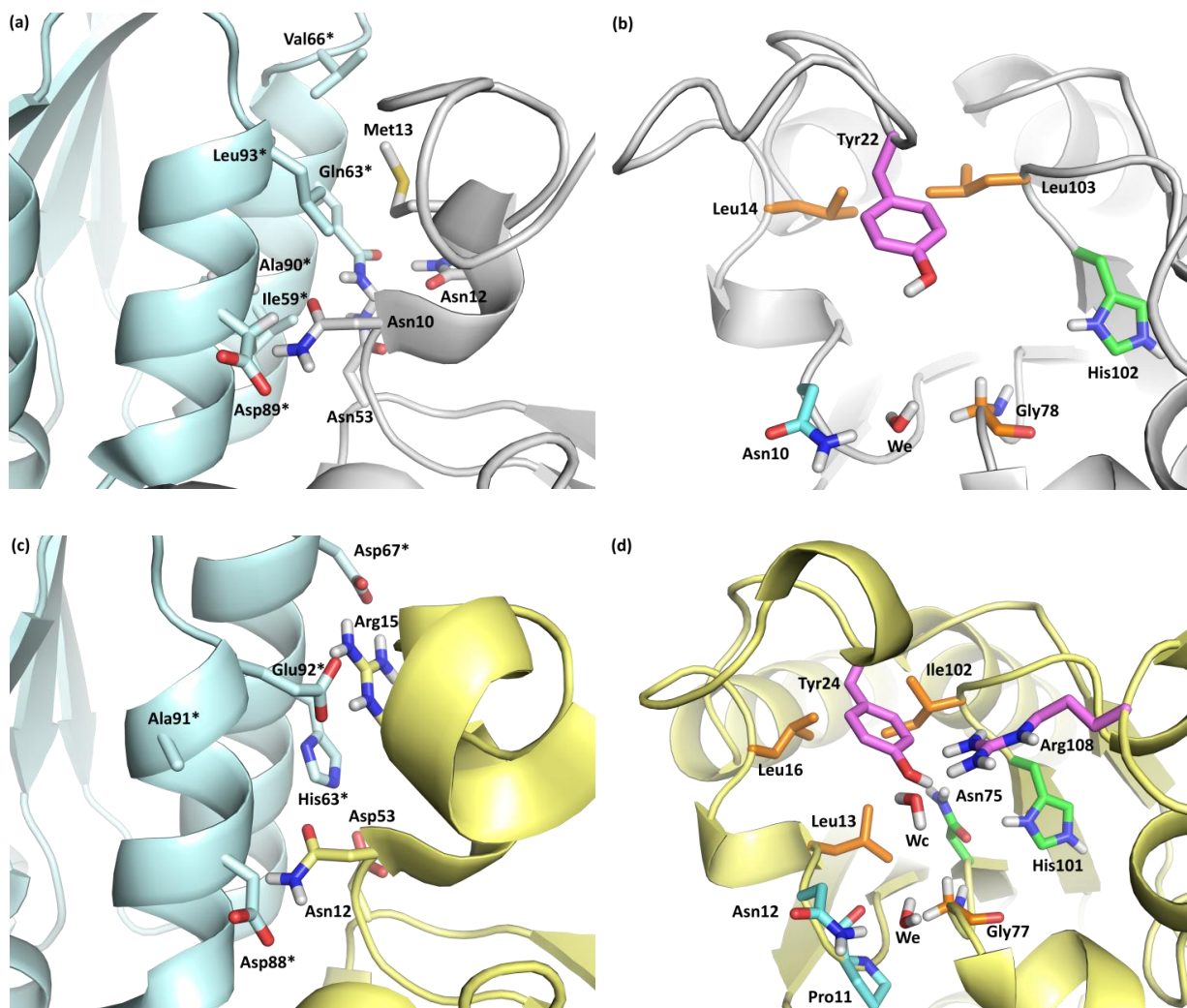


Figure 4. A detailed view of the key residues that contribute most significantly to explaining the differences in inhibitory potency, as revealed by the COMBINE analysis: DHQ2-Hp (a and b) and DHQ2-Mt (c and d). Figures a and c show the interface pocket between two symmetry-related neighboring chains close to the enzyme active site. Note how there are important differences regarding the type of residues lining this pocket in both enzymes. Side chains of relevant residues are shown and labeled.

## Conclusions and Final Remarks

Two COMBINE analyses have been conducted to quantify the contribution of each active site residue to inhibition of the type II dehydroquinases from *H. pylori* and *M. tuberculosis* by a large series of compounds spanning more than four orders of magnitude in inhibition constants. The best models attained acceptable figures of merit, as assessed by correlation

coefficients,  $r^2$ , of 0.83 and 0.88 and cross-validated correlation coefficients,  $q^2$ , of 0.73 and 0.71, respectively.

The QSAR model for DHQ2-Hp identifies the van der Waals interactions with Asn10, Met13, Leu14, Tyr22, Leu103, His102 and Gly78, and Gln63\* and Ala90\* of a symmetry-related neighboring subunit, as well as the electrostatic and hydrogen bonding interactions with the essential water molecule, as those that contribute more significantly to explaining the differences in inhibitory potency. For DHQ2-Mt, the model is dominated by van



der Waals interactions with active site residues Tyr24, Leu13, Leu16, Ile102, Gly77, Asn75 and His101, and with Ala91\* and Glu92\* of the neighboring subunit, as well as by electrostatic and/or by hydrogen bonding interactions with the essential water molecule and the crystallographic water molecule close to Arg108 and Tyr24.

It is noteworthy that, for both enzymes, the COMBINE highlight the importance of the water molecule involved in the enzymatic mechanism, which fixes the natural substrate and the intermediate in the active site. A second key contribution in the inhibition activity is the C2 binding pocket, particularly for DHQ2-Mt, which involves residues Leu13, Leu16, Ile102 and Gly77. In addition, significant differences between both enzymes are detected in the way the inhibitors interact with the essential tyrosine, the residue that initiates enzymatic catalysis. We consider that these differences arise from differences in loop flexibility and motion between the two enzymes.

Finally, the COMBINE approach has allowed us to quantify the importance of ligand interactions at the interface pocket close to the active site, to highlight the key differences between both enzymes and to provide guidance for further inhibitor design. Thus, for DHQ2-Hp, enzyme inhibition might be enhanced by promoting interactions between the ligand and residues Gln63\* and Ala90\*, which are located at the “bottom” of the interface, and avoiding contacts with residues Asp89\* and Leu93\*, which are located at the “upper part” of the interface. On the other hand, for DHQ2-Mt, improving the van der Waals interactions with Glu92\*, a residue located at the entrance to the interface, might provide more potent inhibitors. For both enzymes, extending the contacts with Asp88\*, which is involved in the recognition of the hydroxyl group at the C4 position in both the natural substrate and the reaction intermediate, would be unfavorable.

## Experimental Section

**Ligand minimization.** Initial ligand geometries were first refined by using the PM6 hamiltonian, as implemented in Gaussian 09,<sup>[40]</sup> and further energy-minimized using a restricted Hartree–Fock (RHF) method and a 6–31G(d) basis set while fixating the three dihedral angles defining the orientation of the carboxylate group and the hydroxyl groups at C1 and C5. The charge distribution for all the ligands was obtained by fitting the electrostatic potential-derived (ESP) atom-centered charges as implemented in Gaussian 09, to a restrained electrostatic potential (RESP)<sup>[41,42]</sup> point-charge model, as implemented in AMBER<sup>[43]</sup> suite of programs using a polarizable continuum model of solvation. The missing bonded and non-bonded parameters were assigned, by analogy or through interpolation from those already present in the AMBER database (GAFF).

**Minimization of crystallographic DHQ2-ligand binary complexes.** These studies were carried out using the enzyme geometries found in several crystal structures of the binary complexes DHQ2-Hp and DHQ2-Mt. For the former enzyme, eight crystal structures were used with accession codes 2WKS,<sup>[7]</sup> 1J2Y,<sup>[35]</sup> 2C4W,<sup>[24]</sup> 2C57,<sup>[24]</sup> 2XB9,<sup>[23]</sup> 2XDA,<sup>[20]</sup> 2C4V<sup>[24]</sup> and 2XD9<sup>[20]</sup>. For the latter, thirteen crystal structures were used with accession codes 1H0R,<sup>[36]</sup> 1H0S,<sup>[36]</sup> 2Y71,<sup>[8]</sup> 2Y76,<sup>[8]</sup> 2Y77,<sup>[8]</sup> 2XB8,<sup>[23]</sup> 3N59,<sup>[9]</sup> 3N8K,<sup>[9]</sup> 3N76,<sup>[9]</sup> 3N87,<sup>[9]</sup> 3N8N,<sup>[9]</sup> 3N7A<sup>[9]</sup> and 1H05<sup>[37]</sup>. Considering that, to guarantee a successful COMBINE analysis, all the proteins employed must have exactly the same number of residues, not solved residues were incorporated from the fully resolved crystal structures, i.e. 2WKS<sup>[7]</sup> for DHQ2-Hp and 2XB8<sup>[23]</sup> for DHQ2-Mt. With the exception of two water molecules located in the active site, all the other water molecules and the cocrystallized molecules originating from the enzyme storage buffer (TRIS, glycerol, sulfate, etc.) were removed from the complexes. This water molecules are the conserved one (WAT<sub>c</sub>) involved in the

enzymatic mechanism and a crystallographic water molecule (WAT<sub>c</sub>) located near the essential tyrosine and the conserved arginine, i.e. Arg109, for DHQ2-Hp, and Arg108, for DHQ2-Mt. First, energy minimization of the crystallographic DHQ2-ligand binary complexes used in this study was performed using the Generalized Born (GB) solvation model (igb = 1) with no cutoff for non-bonded interactions (cut = 999), no periodic boundary conditions (ntb = 0) and the steepest descent algorithm. Taking into account that unfolding and refolding studies on DHQ2 have shown that the trimer<sup>[44]</sup> is the biological unit of the enzyme, the trimer was used for these studies. Hydrogens were added to the protein using the web-based PROPKA3.1 server,<sup>[45–48]</sup> which assigned protonation states to all titratable residues at the chosen pH of 7.0. However,  $\delta$  and/or  $\epsilon$  protonation was manually corrected for His101/His102 (dual) of the active site due to mechanistic considerations and the fact that PROPKA does not consider any bound ligand in the calculation. Molecular mechanics parameters from the ff03 and GAFF force fields were assigned to the protein and waters and to the ligand, respectively. Energy minimizations were performed using the Sander program of AMBER 10. The positions of hydrogens were first optimized in vacuo by running 500 steps of steepest descent followed by 500 steps of conjugate gradient energy minimization. Protein side-chain optimization then followed by performing 1000 steps of steepest descent followed by 1000 steps of conjugate gradient energy minimization. A force constant of 50 kcal mol<sup>-1</sup> Å<sup>-2</sup> was used for positionally restraining all the atoms not subjected to energy refinement. Finally, the entire complex was adapted to the force field by performing 1000 steps of steepest descent followed by 1000 steps of conjugate gradient energy minimization.

**Minimization of the DHQ2-ligand complexes.** Bound ligands in the energy-minimized crystallographic binary complexes were replaced by the set of congeneric inhibitors (supporting information) and aligned. Energy optimization of each complex was achieved by using the steepest descent algorithm and the GB implicit solvent model until the root-mean-square of the potential energy gradient was below 0.05 kcal.mol<sup>-1</sup> Å<sup>-1</sup>.

**COMBINE analysis.** AMBER coordinate and topology files for the set of refined ligand-enzyme complexes containing two water molecules as explained above, were directly used as input to the gCOMBINE program.<sup>[49]</sup> The inhibitory activity data for the set of inhibitors that are available as  $K_i$  values (expressed in molar units, supporting information), were converted to negative logarithmic values (i.e., p*K<sub>i</sub>*), which range from 3.4 to 7.0 for DHQ2-Hp and from 3.7 to 7.6 for DHQ2-Mt. The interaction energy matrix was then subjected to partial least squares (PLS) analysis using the p*K<sub>i</sub>* values as the experimental observables, the leave-one-out method for cross-validation and a uniform dielectric constant of 4.0 to calculate the electrostatic interactions. Any positive van der Waals energy value was truncated to 0.0 kcal mol<sup>-1</sup> and any variable with a standard deviation below 0.1 kcal mol<sup>-1</sup> within the series was removed. The quality of the models was assessed by the correlation coefficient ( $r^2$ ) and the cross-validated correlation coefficient ( $q^2$ ).

Figures were prepared using PYMOL<sup>[50]</sup> and CHIMERA<sup>[51]</sup>.

## Acknowledgements

*Financial support from the Xunta de Galicia (10PXIB2200122PR and GRC2010/12 to C.G.-B.) and Comunidad de Madrid (S2010-BMD-2457 to F.G.) and the Spanish Ministry of Science and Innovation (SAF2010-15076 to C.G.-B. and SAF2009-13914-C02-02 to F.G.) is gratefully acknowledged. A.P. and C.C. thank the Spanish Ministry of Science and Innovation for their respective FPU fellowships. We are also grateful to the Centro de*

**Keywords:** binding interactions, dehydroquinase, structure-activity relationship, competitive inhibitors, antibiotics

- [1] C. Abell, Enzymology and molecular biology of the shikimate pathway. In *Comprehensive Natural Products Chemistry*, Sankawa, U., Ed.; Elsevier Science Ltd.: Oxford, 1999. pp 573–607.
- [2] a) F. Roberts, C. W. Roberts, J. J. Johnson, D. E. Kyle, T. Krell, J. R. Coggins, G. H. Coombs, W. K. Milhous, S. Tzipori, D. J. Ferguson, D. Chakrabarti, R. McLeod, *Nature* **1998**, 393, 801–805; b) P. J. Keeling, J. D. Palmer, R. G. Donald, D. S. Roos, R. F. Waller, G. I. McFadden *Nature* **1999**, 397, 219–220.
- [3] Essential genes data base: [www.essentialgene.org](http://www.essentialgene.org).
- [4] J. Harris, C. González-Bello, C. Kleanthous, J. R. Coggins, A. R. Hawkins, C. Abell, *Biochem. J.* **1996**, 319, 333–336.
- [5] A. W. Roszak, D. A. Robinson, T. Krell, I. S. Hunter, M. Frederickson, C. Abell, J. R. Coggins, A. J. Laphorn, *Structure* **2002**, 10, 493–503.
- [6] V. F. V. Prazeres, L. Castedo, H. Lamb, A. R. Hawkins, C. González-Bello, *ChemMedChem* **2009**, 4, 1980–1984.
- [7] V. F. V. Prazeres, L. Tizón, J. M. Otero, P. Guardado-Calvo, A. L. Llamas-Saiz, M. J. van Raaij, L. Castedo, H. Lamb, A. R. Hawkins, C. González-Bello, *J. Med. Chem.* **2010**, 53, 191–200.
- [8] L. Tizón, J. M. Otero, V. F. V. Prazeres, A. L. Llamas-Saiz, G. C. Fox, M. J. van Raaij, H. Lamb, A. R. Hawkins, J. A. Ainsa, L. Castedo, C. González-Bello, *J. Med. Chem.* **2011**, 54, 6063–6084.
- [9] M. V. Dias, W. C. Snee, K. M. Bromfield, R. J. Payne, S. K. Palaninathan, A. Ciulli, N. I. Howard, C. Abell, J. C. Sacchettini, T. L. Blundell, *Biochem. J.* **2011**, 436, 729–739.
- [10] C. Sánchez-Sixto, V. F. V. Prazeres, L. Castedo, H. Lamb, A. R. Hawkins, C. González-Bello, *J. Med. Chem.* **2005**, 48, 4871–4881.
- [11] C. Sánchez-Sixto, V. F. V. Prazeres, L. Castedo, S. W. Suh, H. Lamb, A. R. Hawkins, F. J. Cañada, J. Jiménez-Barbero, C. González-Bello, *ChemMedChem* **2008**, 3, 756–770.
- [12] V. F. V. Prazeres, C. Sánchez-Sixto, L. Castedo, H. Lamb, A. R. Hawkins, A. Riboldi-Tunnicliffe, J. R. Coggins, A. J. Laphorn, C. González-Bello, *ChemMedChem* **2007**, 2, 194–207.
- [13] M. D. Toscano, R. J. Payne, A. Chiba, O. Kerbarh, C. Abell, *ChemMedChem* **2007**, 2, 101–112.
- [14] R. J. Payne, A. Riboldi-Tunnicliffe, O. Kerbarh, A. D. Abell, A. J. Laphorn, C. Abell, *ChemMedChem* **2007**, 2, 1010–1013.
- [15] R. J. Payne, F. Peyrot, O. Kerbarh, A. D. Abell, C. Abell, *ChemMedChem* **2007**, 2, 1015–1029.
- [16] a) A. T. Tran, K. M. Cergol, N. P. West, E. J. Randall, W. J. Britton, S. A. Bokhari, M. Ibrahim, A. J. Laphorn, R. J. Payne, *ChemMedChem* **2011**, 6, 262–265; b) A. T. Tran, K. M. Cergol, W. J. Britton, S. A. I. Bokhari, M. Ibrahim, A. J. Laphorn, R. J. Payne, *Med. Chem. Commun.* **2010**, 1, 271–275.
- [17] B. Blanco, A. Sedes, A. Peón, H. Lamb, A. R. Hawkins, L. Castedo, C. González-Bello, *Org. Biomol. Chem.* **2012**, 10, 3662–3676.
- [18] M. Frederickson, E. J. Parker, A. R. Hawkins, J. R. Coggins, C. Abell, *J. Org. Chem.* **1999**, 64, 2612–2613.
- [19] M. Frederickson, A. W. Roszak, J. R. Coggins, A. J. Laphorn, C. Abell, *Org. Biomol. Chem.* **2004**, 2, 1592–1596.
- [20] S. Paz, L. Tizón, J. M. Otero, A. L. Llamas-Saiz, G. C. Fox, M. J. van Raaij, H. Lamb, A. R. Hawkins, A. J. Laphorn, L. Castedo, C. González-Bello, *ChemMedChem* **2011**, 6, 266–272.
- [21] A. T. Tran, N. P. West, W. J. Britton, R. J. Payne, *ChemMedChem* **2012**, 7, 1031–1043.
- [22] M. F. Schmidt, O. Korb, N. I. Howard, M. V. B. Dias, T. L. Blundell, C. Abell, *ChemMedChem* **2013**, 8, 54–58.
- [23] A. Peón, J. M. Otero, L. Tizón, V. F. V. Prazeres, A. L. Llamas-Saiz, G. C. Fox, M. J. van Raaij, H. Lamb, A. R. Hawkins, F. Gago, L. Castedo, C. González-Bello, *ChemMedChem* **2010**, 5, 1726–1733.
- [24] D. A. Robinson, K. A. Stewart, N. C. Price, P. A. Chalk, J. R. Coggins, A. J. Laphorn, *J. Med. Chem.* **2006**, 49, 1282–1290.
- [25] L. Lence, L. Tizón, J. M. Otero, A. Peón, V. F. V. Prazeres, A. L. Llamas-Saiz, G. C. Fox, M. J. van Raaij, H. Lamb, A. R. Hawkins, C. González-Bello, *ACS Chem. Biol.* **2012**, ASAP, doi:10.1021/cb300493s.
- [26] A. R. Ortiz, M. T. Pisabarro, F. Gago, R. C. Wade, *J. Med. Chem.* **1995**, 38, 2681–2691.
- [27] A. R. Ortiz, M. Pastor, A. Palomer, G. Cruciani, F. Gago, R. C. Wade, *J. Med. Chem.* **1997**, 40, 1136–1148.
- [28] C. Pérez, M. Pastor, A. R. Ortiz, F. Gago, *J. Med. Chem.* **1998**, 41, 836–852.
- [29] T. Wang, R. C. Wade, *J. Med. Chem.* **2001**, 44, 961–971.
- [30] J. Kmunicek, S. Luengo, F. Gago, A. R. Ortiz, R. C. Wade, J. Damborsky, *Biochemistry* **2001**, 40, 8905–8917.
- [31] C. Coderch, A. Morreale, F. Gago, *Anticancer Agents Med. Chem.* **2012**, 12, 219–225.
- [32] T. Wang, S. Tomic, R. R. Gabdouliline, R. C. Wade, *Biophys. J.* **2004**, 87, 1618–1630.
- [33] S. Tomic, L. Nilsson, R. C. Wade, *J. Med. Chem.* **2000**, 43, 1780–179.
- [34] T. Wang, R. C. Wade, *J. Med. Chem.* **2002**, 45, 4828–4837.
- [35] B. I. Lee, J. E. Kwak, S. W. Suh, *Proteins* **2003**, 51, 616–617.
- [36] Protein data bank by D. A. Robinson, A. W. Roszak, M. Frederickson, C. Abell, J. R. Coggins, A. J. Laphorn, to be publish.
- [37] L. D. Evans, A. W. Roszak, L. J. Noble, D. A. Robinson, P. A. Chalk, J. L. Matthews, J. R. Coggins, N. C. Price, A. J. Laphorn, *FEBS Lett.* **2002**, 530, 24–30.
- [38] T. Krell, A. R. Pitt, J. R. Coggins, *FEBS Lett.* **1995**, 360, 93–96.
- [39] T. Krell, M. J. Horsburgh, A. Cooper, S. M. Kelly, J. R. Coggins, *J. Biol. Chem.* **1996**, 271, 24492–24497.
- [40] Gaussian 09, Revision A.2, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.
- [41] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, 117, 5179–5197.
- [42] <http://q4md-forcedfieldtools.org/RED/resp/>
- [43] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, O. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, 26, 1668–1688.
- [44] N. C. Price, D. J. Boam, S. M. Kelly, D. Duncan, T. Krell, D. G. Gourley, J. R. Coggins, R. Virden, A. R. Hawkins, *Biochem. J.* **1999**, 338, 195–202.
- [45] H. Li, A. D. Robertson, J. H. Jensen, *Proteins* **2005**, 61, 704–721.
- [46] D. C. Bas, D. M. Rogers, J. H. Jensen, *Proteins* **2008**, 73, 765–783.
- [47] M. H. M. Olsson, C. R. Søndergard, M. Rostkowski, J. H. Jensen, *J. Chem. Theor. Comp.* **2011**, 7, 525–537.
- [48] C. R. Søndergard, M. H. M. Olsson, M. Rostkowski, J. H. Jensen, *J. Chem. Theor. Comp.* **2011**, 7, 2284–2295.
- [49] R. Gil-Redondo, J. Klett, F. Gago, A. Morreale, *Proteins* **2010**, 78, 162–172.
- [50] W. L. DeLano, The PyMOL Molecular Graphics System. (2008) DeLano Scientific LLC, Palo Alto, CA, USA. <http://www.pymol.org>.
- [51] <http://www.cgl.ucsf.edu/chimera>

Received: ((will be filled in by the editorial staff))

Published online: ((will be filled in by the editorial staff))



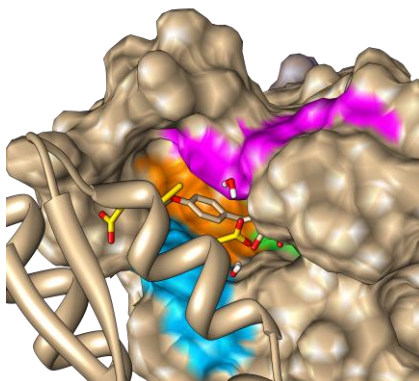
## Entry for the Table of Contents

### FULL PAPERS

---

#### **Interactions driving inhibition:**

Quantitative-activity relationship studies conducted with type II dehydroquinase inhibitors from *H. pylori* and *M. tuberculosis* supports the importance of the hydrogen bonding interaction of the inhibitors with the water molecule involved in the enzymatic mechanism and identified important differences in the ligand interactions with the interface pocket close to the active site for both enzymes.



*Antonio Peón, Claire Coderch, Federico Gago and Concepción González-Bello\**

**Page No. – Page No.**

**Comparative Binding Energy (COMBINE) Analysis for Understanding the Binding Determinants of Type II Dehydroquinase Inhibitors**